iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** III **Month of publication:** March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3302

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

A Survey on Different Machine Learning Approaches

Vrushabh Shende[1], Pratik Agrawal[2], Komal Hole[3]
[1, 2, 3]*Computer Science and Engineering, PRMIT&R College Badnera, Amravati, India*

Abstract: ***Today's infrastructure technology contains much more data than early days. This data is complex in structure and massive in size. Number of different aspects and fields uses this huge complex data i.e. developing new innovations, making perfect decisions, predicting the output accurately. These purposes can be carried out by using AI (Artificial Intelligence). To understand AI, understanding Machine Learning is must, as machine learning is the base of AI. Machine learning is helpful for decision making and predicting accurate and precise output. System learns directly from examples, experiences and data with the help of machine learning. Analyzation and prediction from the available vital data can be carried out with the help of machine learning. Machine learning can be differentiating in several approaches based on working procedure and their behavior. There are several types of machine learning methodologies which are classified depending upon the data such as input and output processed by them and their behavior. These methodologies and machine learning approaches have been adapted by various types of applications. The focus point of this analysis i.e. research paper is to identify the better machine learning approach, which can lead to accurate learning with high positive rate. This analysis work also shows the merits and demerits of earlier machine learning techniques.***

Keywords: Machine learning, artificial intelligence, machine learning approaches

I. INTRODUCTION

In this era, every infrastructure that uses technology is trying to use artificial intelligence to reduce the work load. As machine learning is the basic of artificial intelligence, it makes artificial intelligence reliable on machine learning. Due to artificial intelligence, machine learning is being used by many technologies. Machine learning is use for developing the ability to learn without being explicitly programmed by the system. Machine learning helps to analyze the data. Machine learning is based on different algorithms which are used to make predictions over data. Due to the large amount of data, the need for machine learning is increasing day-by-day. To predict precise and accurate output, ML uses different algorithms, it also helps to learn. The creativeness in the system is done with the help of algorithm. Explicitly programming is not needed. Machine learning keeps the record of the previous examples, observation over data used as input and the generate output, instructions used in the program, or some direct experiences. These records helps machine learning to learn new things about the output generation. So that output can be more precise and accurate in terms of previous data. Machine leaning follows the patterns drawn by the algorithm on the basis of previous example. The main goal of this whole machine learning is to avoid human interventions or assistance and allow the system to learn automatically and adjust actions accordingly.

II. HISTORY

A. The first known example can be seen in 1950. It was a test known as "Turing Test", Developed by Alan Turing. The purpose of the test was to determine computer's real intelligence. For that the system needs to fool a human into believing it is also a human.

B. The second example comes in 1952. Arthur Samuel developed the first machine learning program. This program is the checkers game. This game was developed by IBM. The game was improved by the computer by studying the winning moves and strategies used by the gamer.

C. In 1957 the first neural network was designed by Frank Rosenblatt. It was also known as the perceptron. The perceptron simulates the thoughts processes by the human brain.

D. A decade later in 1967, an algorithm was written known as "nearest neighbor". This algorithm helps in using basic pattern recognition.

E. Some Stanford University students invented the "Stanford Cart" in 1979. The main feature of this car was that this car can navigate obstacles in the room on its own.

F. The concept of Explanation Based Learning (EBL) was introduced in 1981 by Gerald Dejong. Using this system can crates its own general rules. For that the computer analyses training data and discard unimportant data.

G. Terry Sejnowski invents NetTalk in the year 1985. NetTalk pronounce the verbal words like a baby does.

H. Program for learning was being developed in 1990 by some scientists. This program helps to analyze the large amount of complex data and can make a conclusion out of it.

I. In 1997, IBM created the Deep Blue system that beats the world champion at chess.

J. In 2006, Geoffrey Hinton introduce a new concept in machine learning i.e. "see". This new concept is known as "deep learning". The main feature of this algorithm is to analyze the video and distinguish text from the video images.

K. The Microsoft Kinect developed a technology in 2010, which can track different 20 human features at a rate of 30 times per second. This technology helps people to interact with the system by the means of movements and gestures.

L. The IBM computer, The Watson beats its human competitor in 2011. The game played along them was Jeopardy. This game is played by two competitor by answering questions in natural language.

M. The project Google Brain was leads by two scientist in 2012. These scientists were Jeff Dean and Andrew Ng (Stanford University). This project is able to detect patterns in videos and images using all of the Google infrastructure capacity and develops a deep neural network.

N. Deep Mind is the British startup that was bought by Google. This startup contains DNN capabilities. The startup simply read the pixels on the screen and play the game.

O. "Distributed Machine Learning Toolkit" was in created in 2015 by Microsoft. This allows for the efficient distribution of machine learning problems to multiple computers

P. In 2016 IBM Watson's natural language processed by The North Face in a mobile app.

III. APPLICATION AREA OF MACHINE LEARNING

Machine learning helps to build artificial intelligence. It is the base of artificial intelligence. The machine learning uses different algorithm methodologies. These algorithms helps the machine learning to analyze large amount of complex data and generate more accurate and precise output. The algorithms finds out the pattern to generate the output. These patterns can be drawn through the previous example. It can be concluded that the machine learning learns from the previous experiences, draw the pattern from it and generate the output. The machine learning technology is being adapted by many technological firms wherever it is needed.

1) *Healthcare*: technology is the essential part in the medical field, as many diagnosis and medical treatments need the use of technology. Machine learning can play the vital role in the field of healthcare. The machine learning can useful for diagnosis improvisation, monitoring treatment plans, curing the disease. The machine learning is use for predicting the output based on inputs and previous experiences can help the doctors to diagnose the most harmful disease. That will leads to better patient care and services. Machine learning will be able to detect any kind of dangerous medical event. Diagnosis can be improve by analyzing huge amount of medical data, draw patterns from it depending upon previous patients and enhance medical treatment and care. This will lead to prevent and type of hazardous diseases.

2) *Retails*: machine learning is very much useful in retailing. Machine learning collects the data from users who purchase the products online. This data is then analyzed by the machine learning. Machine learning algorithm captures the data. This helps the ML to point out the dots. By connecting these dots, ML makes draws a pattern for shopping. This pattern helps to discover the similarities and differences of the shopping by customers. It helps the site to add the necessary products for the customers. As an example, online shopping sites can be considered. These sites study the shopping behavior of the customer. The suggestions are then shown to the customers based on the price range, coupons, type of product, brands.

3) *Financial Services*: financial services like banks and financial institutes can use the previous data. Many financial sectors can avoid dangerous events like frauds, bankrupts', intruder stealing. The machine learning can analyze the data of the bank and the employee data and keep the track of the flow. So if anything malicious problem happens, machine learning can suggest the way to avoid it. Machine learning leans the data and behavior from the database and human interaction. And this data is use for more informed decision-making. The ML can also suggest investment strategies to avoid any loss on the basis of current market status and previous market values.

4) *Automotive*: as like any other industry, automotive are trying to develop themselves. So that they will be able to stand in this competitive world. They are making new inventions to increase the market rate of the firm. The main feature for of machine learning i.e. to learn from data and example can helps the automotive industry. This operations can be improved with the help of ML. machine learning can record the customers experience before, during, and after purchase. With the help of this data the automotive industries, can increase their productivity so that the customers can get satisfied. Machine learning can also analyze the impact of previous marketing efforts. Based on these examples and data ML can make new investment strategies.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

IV. MACHINE LEARNING APPROACHES

As discussed earlier machine learning has different types of approaches. Machine learning needs different approaches to deal with different types of data. There are three categories where machine learning approaches are classified.

1) *Supervised*: in supervised machine learning approach the data is labeled, for instance input and output. This approach takes the data and predicts the outcome on the given set of example and equations. Using these labels and data, this approach draws a unique pattern. Using that pattern the approach tries to predict the next outcome. Every time the example is completely solved, the approach records it and makes more accurate and precise output.
2) *Unsupervised*: as the opposite of supervised approach, unsupervised approach don't use the labels on data. To make the complex data simple and easy, this approach divides and organizes the data into groups that are known as clusters. These clusters help to understand and describe the structure of complex data.
3) *Reinforcement*: This approach doesn't follow any pattern. This approach simply choose any action any make the output. These outputs make a data points. Later the approach checks and learns how good the decision was. Over time, the approach changes its strategy to learn better and achieve the better output.

Following are shown different types of machine learning algorithm.

a) *Linear Regression:* Linear regression follows the supervised approach. It is very much popular algorithm due to its simplicity. To make the predictions over output, this algorithm needs continuous variables. As it is supervised approach, the data variables are labeled as dependent variable and independent variable. To predict the future outcome, this algorithm attempts to make relationship between them by using the regression line. The regression line helps to minimize the distance between the data points. The linear regression equation makes a straight slope line. The can equation is

$$Y = a*X + b$$

Where, Y is the dependent variable, X is the independent variable, 'a' is the slope, and b is the intercept.

The linear regression can be divided in two types:

Simple Linear Regression – Where there is only one independent variable.

Multiple Linear Regression – Where there is more than one independent variable.

b) *Logistic Regression:* Logistic regression follows the unsupervised approach. In logistic regression the data is not labialized. This algorithm follows only discrete values, as yes(1) or no(0). This algorithm drives the output depending upon the program or Logistic Functions. Using this function a relationship can be determine between dependent and independent variables. On graph, this function makes a S-shaped curve. This is opposite to linear regression as it makes a straight line. This line divides the data in the discrete values that will be either yes(1) or no(0). So this algorithm is applicable if the data outcome can be predicted in binary i.e. 0 or 1. Other than that, different algorithms can be used.

c) *Decision Tree:* This is the simplest algorithm of machine learning. Decision tree follows the supervised approach. It is also easiest to understand. The main goal for decision tree is to perform classification. As it is supervised algorithm, it labels the data. The decision tree can also form regression tree. Decision tree divides the data into small group which feels to be same but are rather different to each other. Decision tree follows many different algorithms. Decision tree needs different types of algorithm for different types of data.

d) *Support Vector Machine:* Support Vector Machine (SVM), is a supervised machine algorithm. But in some cases it performs a diverse role as an unsupervised algorithm. This algorithm mainly believes in classification, but also useful for regression problems. Like unsupervised algorithm, SVM mainly classifies the data into one of two categories. SVM focuses on number of aspects of differentiate the data. SVM is mostly useful for doing classification. SVM is the hyperplane that divides or "classifies" the data points the best.

e) *K-Nearest Neighbor:* As same like SVM, K-Nearest Neighbour also follows both supervised and unsupervised machine learning approach. This algorithm works for regression as well as classification problems, but more tends towards classification problems. The classification is made on the base on the neighbor data element. In KNN, the nearest data points i.e. neighbor points are use as the value of 'K'. the value of 'K' is not fix for each problem and that is the critical part for this algorithm. As if the value for 'K' is small, then results may contain lot of noise and as a result output won't be accurate. And selecting the higher value destroys the meaning of the algorithm. So to overcome this problem the value of 'K' is set as the square root to the value of total dataset samples.

TABLE I
Comparision Between Different Machine Learning Approaches

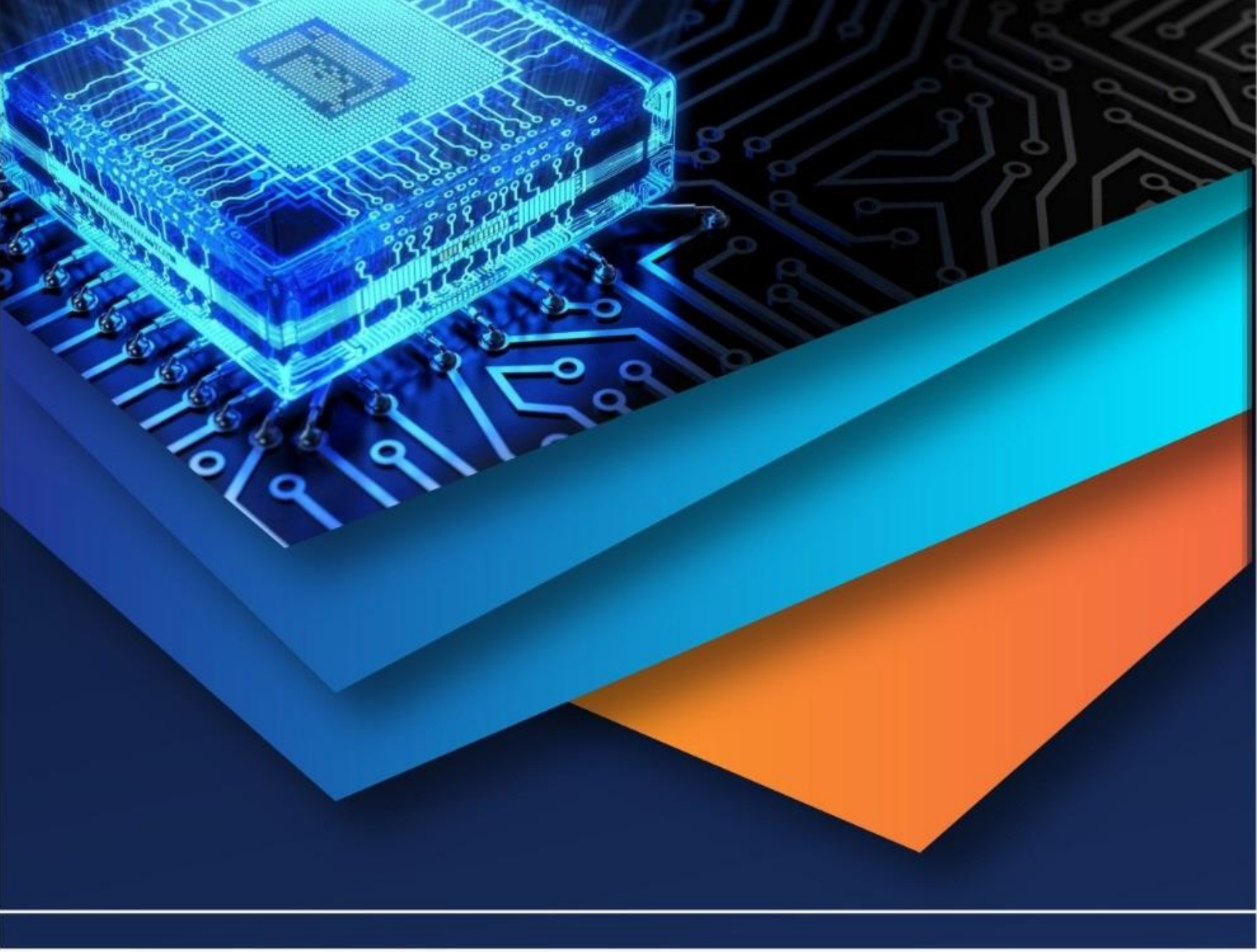
Algorithm	Merits	Demerits	Application
Linear Regression	Easy to explain because it interpreteble. It is easy to use because it requires minimal tuning. Runs fast.	Limited to Linear Relationships Only Looks at the Mean of the Dependent Variable Data Must Be Independent	Estimating Sales Risk Assessment
Logistic Regression	The computational difficulties is decreased High signal quality	The highest amount of data cannot be examined accurately owing to the difficult structure of the procedures	Image Segmentation and Categorization Handwriting recognition Geographic Image Processing
Decision Tree	Easy to understand and instinctual. Decision tree can draw by non-technical people. Making assumption on Non-linearly related data.	Number of decision in tree is inversely proportional to the accuracy. Not useful for continuous variables and result in instability.	In finance for option pricing. Used by banks to classify loan applicants by their probability of defaulting payments.
Support Vector Machine	Offers best performance for classification type problem. Strong assumption cannot be made in this algorithm. It does not over-fit the data.	Hard to choose the good kernel function. Long training time on large data sets	Stock market forecasting by various financial institutions. Breast Cancer Diagnosis
K-Nearest Neighbour	High precision Better prediction rate	Determining the exact results is a complex process and it is labeled data and it does not calculate by KNN methodology owing to its nearest distance relations	Concept Search: If you're searching for semantically similar documents Recommender Systems: If you know a user likes a particular item, then you can recommend similar items for them

V. CONCLUSION

From this analysis work, the conclusion can be made that the technology for today needs to use the Machine Learning. As discussed earlier, many technical industries can be improvised by using the features and different approaches of Machine Learning. To analyze this huge amount of data, which is very much complex in structure, machine learning algorithms will be helpful. As machine learning tends to learns from previous example and tries to make an accurate output every time, this can helps to grow the artificial intelligence. Many algorithms have been studied in this analysis work, so making any specific algorithm a best will be inappropriate, as different algorithm works differently of different type of data. It will be efficient if more researches has been conducted based on reinforcement learning algorithms.

REFERENCE

[1] Frank, E., Hall, M.A. and Pal, C.J. Data Mining:Practical machine learning tools and techniques. Morgan Kaufmann, 2016.
[2] A. KELEMEN, Y. LIANG, S. FRANKLIN "A Comparative Study of Different Machine Learning Approaches for Decision Making" The University of Memphis2014.
[3] R. Reddy,S. Vani, "An Review of Machine Learning Approaches in Data Sensitive Real World Applications" Jour of Adv Research in Dynamical & Control Systems, Vol. 9, No. 3, 2017.
[4] I. MUHAMMAD AND Z. YAN "SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY" 10.21917/ijsc.2015.0133
[5] S. Cates, S. Lawrence, C. Penedo "A MACHINE LEARNING APPROACH TO RESEARCH CURATION FOR INVESTMENT PROCESS" Journal Of Investment Management, Vol. 15, No. 1, (2017), pp. 39–49
[6] B. Uma Maheswari, V. Steffi Joy "MACHINE LEARNING APPROACH FOR IDENTIFYING AND DIAGNOSING THE CANCER TYPES" International Journal of Advanced Engineering Technology E-ISSN 0976-3945
[7] J. Qiu, Q. Wu, G. Ding, Y. Xu and S. Feng "A survey of machine learning for big data processing" Qiu et al. EURASIP Journal on Advances in Signal Processing (2016) 2016:67
[8] S. Danso, E. Atwell and O. Johnson "A Comparative Study of Machine Learning Methods for Verbal Autopsy Text Classification"2014



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY